

## 本周周报(6.3-6.9):

解聪

### 本周工作:

1.对网店的商业交易数据进行清洗。特别地,提取了所有买家留言进行初步统计。

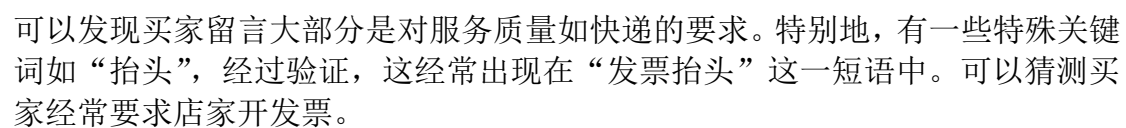
1.使用中科院的 ictclas 工具包进行中文分词,并进行了词频统计。在 ictclas 的基础词库中补充了一些词汇如“圆通”,“韵达”等专有名词。

2.提取了出现频率较高的一些词如下:

一定,641,0	巧克力,447,0	确保,111,0
一点,357,3	希望,664,0	礼品,194,0
不能,144,0	帮忙,175,0	礼物,293,0
不要,1083,0	幸福,169,0	礼盒,318,0
东西,149,0	店主,103,0	祝福,121,0
严实,143,0	店家,178,0	空白,137,0
亲爱,354,0	开心,209,0	红色,177,0
仔细,367,0	快乐,1519,0	结实,151,0
你好,130,0	快递,3444,11	老公,130,0
使用,112,0	感谢,141,0	老婆,216,0
保证,404,0	我们,213,0	老板,237,0
其他,170,0	或者,246,0	联系,365,33
内容,866,0	抬头,160,0	节日,120,0
务必,225,0	拒绝,113,0	袋子,140,0
包装,2632,101	收到,141,0	记得,186,0
卡片,1529,0	新鲜,173,0	谢谢,4044,0
压坏,141,0	日期,423,0	质量,478,0
发票,356,0	时候,100,0	购买,251,4
发货,1673,0	时间,110,0	贺卡,702,0
可以,388,0	最新,112,0	运费,128,0
因为,178,0	有限公司,187,0	还有,119,0
圆通,980,0	朋友,115,123	这个,137,0
地址,167,0	检查,446,0	送给,114,0
天天,165,0	正品,169,0	送货,158,10
好好,109,0	永远,248,0	速度,132,0
如果,177,0	注意,322,0	邮政,104,0
完好,157,0	生产,261,0	邮费,109,0
完整,121,8	生日,503,14	需要,195,0
宝贝,182,0	电话,451,22	顺丰,913,0
尽快,846,0	留言,269,0	顺风,220,1
尽量,183,0	盒子,217,0	食品,161,1
工作日,103,0	破损,128,0	麻烦,1267,1

3.使用 google 翻译将这些词转化为英文。对其中个别词的翻译进行了调整。

4.使用 wordle 在线的文本云生成工具，对这些词进行了可视化，效果如下



### 评审的主要意见

1. 每个评审都提到了“presentation of the paper and the writing are simple confusing”类似观点。表达了他们“hard to follow”文章的意思，尤其是出现在KnotLine的数据聚合与设计那一章。的确写的时候也是这一章最麻烦的。
2. 每个评审都提到了可视设计，虽然很创新，但是文章中缺少了
  - 1) .设计合理性的讨论，设计原则；
  - 2) .用户对设计实用性的反馈，使得看起来太随意，不严谨。大部分评审觉得可视设计太复杂，解读起来费力。  
(也有可能是我们没把设计说清楚，造成了理解困难。总之这部分问题太多了)
3. 案例对用户的研究与评估不够充分。  
没有具体说明分析师的背景。  
只在最后简单罗列了一下用户观点，没有user study。

4. 文章说的不够清晰仔细，比如对于 tweet 的案例，saliency 到底是怎么算出来的;TOS 图中的 1D 的颜色条怎么与其他试图对应的等，情感分析具体怎么做的。  
文章说的不严谨，比如评审认为：我们使用“intuitive”来描述我们的界面，缺少了足够证据。
5. 问题的说法名词的定义有点问题。比如“multi-user behavior”或者“activity class”。一部分程度上也导致了评审理解不清楚。
6. 系统应用场景不明确，即便在案例中也没说具体清楚要做什么。  
应用场景局限，评审觉得有用的可能只有虚假交易的场景。
7. 其他问题。个别评审还有许多有价值的观点，需要进一步分析。

### 我从评审反馈中得出的一些结论：

1. 感觉评审是以严格的 Info Vis 的眼光在审，70%问题是在说设计。说明我们以投 vast 就不用太多笔墨解释设计的想法是错误的。这是我们最大的失误：做了一个相当复杂的可视设计，居然不写相关的设计原则，用户研究等。
2. 两个数据（Twitter，Transaction）存在根本差异，导致我们力求概括统一，导致：
  - 1) .评审看不懂对两种数据统一抽象的描述；
  - 2) .问题的需求与应用场景没法精确定位。黄芯芯和我说过 twitter 不适合，夏菁也提到过，但是的确也很难找到第二个量身定做的数据。说到底可视化为数据服务，而不是数据为可视化服务。
3. 不够实用。我们自己跟别人介绍演示的时候基本大部分人也很难看懂；有一位评审指出我们的系统的功能事实上实用简单的 SQL 语句也能够完成。其实这一点上公司的一些想法，以及对我们提出的一些意见还是值得我们学习的。

### 大部分评审赞同的部分：

1. 可视设计是“novel”以及“interesting”的。（但不严谨也很难懂）；
2. 交易数据的案例是“convincing”的。（twitter 数据的案例的内容评审没提，可能他们连设计都没看懂）；
4. 没有评审对决策树，特征定义的做法提出异议（可能是因为设计问题太多了所以没有说这部分的问题）；
5. 评审并不在意是“Knot”还是“music”，大部分认为我们就是音符式可视化。并没有提出我们预想的“为什么选择音符”这样的问题。

正如评审说的，无论是音符的想法还是虚假交易的案例都是有潜力的。但是我们的工作的确缺陷很多，必须把这些问题一一解决。

### 下一步计划：

1. 完善可视设计与案例的用户研究。
2. 寻找更加统一的可视化设计，包括：
  - 1) 先尝试修改 twitter 的方案；
  - 2) 如果 1) 不是很理想，寻找适合的数据以代替 Twitter 数据。
  - 3) 如果 2) 找不到，考虑是不是只是用于交易数据，但是扩大其应用场景，即寻找除促销，虚假交易外的其他案例。
3. 简化可视设计，降低学习成本。
4. 重新评估系统界面与交互。做到：
  - 1) 实用，人性化。即便没有真正应用到现实中，也要让系统看起来真的具有实用性。
  - 2) 简单，删去不必要的部分与功能。

### 下周工作：

1. 按照计划修改音符项目；
2. 网店数据进一步探索；
3. 争取可以解决这一项目的问题。